International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)

http://www.ijcsjournal.com Volume 3, Issue 1, No 2, 2015. ISSN: 2348-6600
Reference ID: IJCS-078 PAGE NO: 441-444.

Bidirectional Long-Short Term Memory Based Recurrent Neural Network for Handwriting Recognition

Tejashri Chougule[#1], Anis Fatima Mulla[##2]

[1]*Research Scholar.*

[2]*Assistant professor, Department of computer science and engineering, Annasaheb Dange College of Engineering and Technology, Ashta, India.*

[#]*teju917@gmail.com*, [##]*mulla.anis@gmail.com.*

Abstract— **Today, BLSTM is widely used in recurrent neural network for speech and handwriting recognition. In this present paper a novel type of recurrent neural network is specifically designed for sequence labeling tasks where the data is hard to segment and contains long-range bidirectional interdependencies. For many tasks, BLSTM is useful to have access to future, as well as past, context. We refer keyword spotting method for handwritten documents. It is derived from a neural network-based system for unconstrained handwriting recognition. As such it performs template-free spotting, i.e., it is not necessary for a keyword to appear in the training set. The keyword spotting is done using a modification of the CTC Token Passing algorithm in conjunction with a recurrent neural network and a statistical n-gram language model instead of a bigram language model.**

Keywords **— Keyword spotting, offline handwriting, document analysis, historical documents, neural network, BLSTM.**

I. INTRODUCTION

OFFLINE handwritten text recognition such as letters, manuscripts, or entire books is one of the most active areas of research in computer science and it is inherently difficult because of the high variability of writing styles [3], [4].

Making handwritten texts available for searching and browsing is of tremendous value. For example, libraries. Another related application is the segmentation of images of historical documents into meaningful regions, which can be improved with keyword spotting. Transcribing the entire text of a handwritten document for searching is not only inefficient as far as computational costs are concerned, but it may also result in poor performance since misrecognized words cannot be found. Therefore, techniques especially designed for the task of keyword spotting have been developed.

In existing system [1], a bigram language model is used which computes the probability $p(w_1|w_2)$ that word 'w_1' is followed by word 'w_2' in a text for each pair of words (w_1,w_2). Due to this, a high workload in the system and therefore, it is considered as a bottleneck of the system. To improve performance of the existing system, it can be modified to statistical n-gram language model which will compare the number of words at a time and will return probability $p(w_1|w_2|w_3|...|w_n)$.

One of the best existing algorithm deals with this problem that is CTC Token Passing algorithm in conjunction with recurrent neural network and a statistical n-gram language model. Therefore, we have to modify CTC algorithm in terms of a statistical n-gram language model. Next, we review related work from this area.

II. RELATED WORK

A. Word-Based Keyword Spotting:

The task of keyword spotting as detecting a word or a phrase in an image was initially proposed in [5] for printed text and a few years later in [6] for handwritten text. The first methods considered single-word images and adopted approaches common in optical character recognition (OCR).They made use of pixel-wise comparison of the query and the test image (or selected parts of it, called zones of interest (ZOI)) or evaluated a global distance value between the two pixel sets. In [8], different pixel-wise gradient matchings are compared and the authors propose an elastic matching procedure.

B. Line-Based Keyword Spotting:

A different scenario is given if the document is segmented into lines only. A DTW-based system that automatically

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 1, No 2, 2015. ISSN: 2348-6600
Reference ID: IJCS-078 PAGE NO: 441-444.

selects keyword candidates in a handwritten text line is described in [9]. For general systems that rely on an automatic segmentation, a method is proposed in [10] that also takes the probability of a correct segmentation into account. The handwriting recognition have become fairly popular recently, especially using Hidden Markov Models (HMM) [11].

C. Document -Based Keyword Spotting:

To work on completely unsegmented pages of text, a system can either include a segmentation step [12] or take a segmentation-free approach. In [13], a codebook of shapes is used to create a compressed version of each document. A keyword search is then done using the stored shape codebook entries. Finally, a common approach to segmentation-free word spotting is to consider the task as an image retrieval tasks for an input shape representing the word image [14].

In this paper, we have focus on line based keyword spotting and a statistical n-gram language model.

III. PROBLEM DEFINATION

We have studied some proposed techniques in related work. But all these techniques incur the problem of producing a high workload on the system. Such a problem degrades the system performance in terms of execution time and space. If system generates a statistical n-gram language model, then higher processing time it consumes. The CTC Token Passing algorithm in conjunction with recurrent neural network and statistical n-gram language model overcomes this limitation. This algorithm mines a likely sequence of words by using keyword spotting technique.

IV. PROPOSED SYSTEM

In this paper, we present a keyword spotting method for handwritten text based on BLSTM Neural Networks. The application of these networks in conjunction with the so-called CTC Token Passing algorithm to produce a transcription of handwritten text was presented in [2]. With our system, fast and reliable keyword spotting can be performed without the need for either transcribing the text line or segmenting it into individual words. *(Keyword spotting refers to the process of retrieving all instances of a given word from a document.)*

In past, no language models of higher order were considered. But here, relatively little work has been done on the language model. Indeed, most systems rely on the same hidden Markov models and a statistical bigram language model that have been used for decades in speech and handwriting recognition, despite their well-known shortcomings. This paper proposes an alternative approach based on a BLSTM neural network with a statistical n-gram language model.

A block diagram of proposed system as follow:

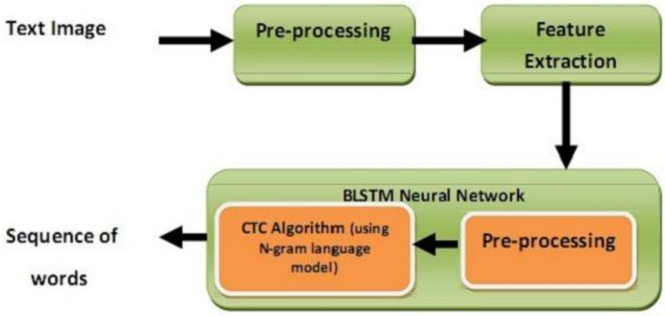

Fig 1. A block diagram.

V. METHODOLOGY

A. *PRE-PROCESSING*

Each person has a different writing style with its own characteristics (see in Fig 2). This fact makes the recognition task complicated. To reduce variations in the handwriting texts, a number of pre-processing operations are applied.

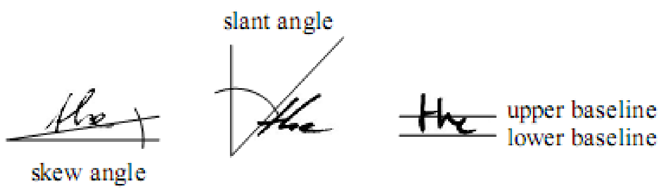

Fig 2 Writing characteristics: Skew angle, slant angle and baselines.

In the presented system the following five pre-processing steps are carried out (for more details see [7]):

1) Line separation: The given page is divided into individual lines.

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)

http://www.ijcsjournal.com Volume 3, Issue 1, No 2, 2015. ISSN: 2348-6600
Reference ID: IJCS-078 PAGE NO: 441-444.

2) Skew Correction: The text line is horizontally aligned. For this purpose the skew angle is need to be determined (see figure 2 left part).
3) Slant Correction: Applying the shear transformation, the writing's slant is transformed into an upright position (see figure 2 middle part).
4) Line Positioning: The text line's total extent in vertical direction is normalized to a standard value. Moreover, applying a vertical scaling operation the location of upper and lower baseline (see figure 2 right part) is set to be standard position.
5) Horizontal Scaling: The variations in the width of the writing are normalized.

After pre-processing next part is feature extraction.

B. *Feature Extraction*

For algorithmic processing, the normalized text line image is represented by a sequence of N feature vectors $x_1 \dots x_n$ with $x_i \; \varepsilon \; IR^n$. This sequence is extracted by a sliding window moving from the left to the right over the image. At each of the N positions of the sliding window, n features are extracted. The sliding window has a width of one pixel. It is moved in steps of one pixel, i.e., N equals the width of the text line. From each window, $n = 9$ geometric features are extracted, three global and six local ones. For further details on the feature extraction step, we refer to [1], [7].

From each line, a sequence of feature vectors is extracted, which is then submitted to the neural network.

C. *BLSTM neural network*

Applying BLSTM NN to handwriting recognition consists of two parts.

1) *Pre-processing by BLSTM NN:* To overcome the vanishing gradient problem that describes the exponential increase or decay of information in recurrent connections in a neural network, the nodes in the hidden layer are replaced by long short-term memory (LSTM) cells, displayed in Figure 4. We refer to [2].

   The gates of these cells are normal nodes and control the flow of information into and out of each cell. When the input gate is open, the central nodes value is replaced by the output activation of the net input node. When the output gate is open, information flows out into the network and when the forget gate is open, the cell's memory is reset to zero.

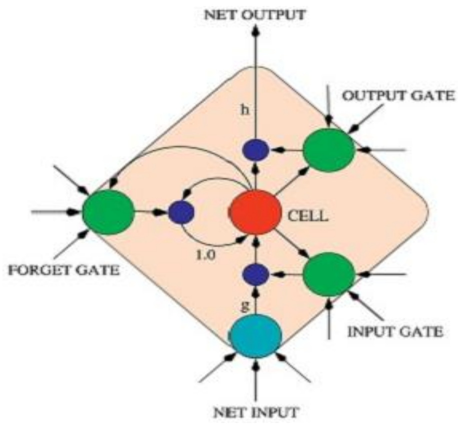

Fig 3. LSTM memory block with one cell.

In pre-processing phase, combining BRNNs and LSTM gives BLSTM [2]. The BRNN is bidirectional RNN, meaning the text line is processed from both left-to-right and right-to-left. This is done because context from both sides of a character is useful to improve the recognition. The information from two separate input layers is collected in two separate LSTM layers, respectively, and finally joined in the output layer. This is illustrated in figure. 5.

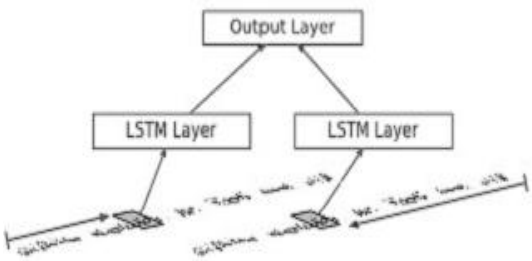

Fig 4. An illustration of the mode of operation of the BLSTM Neural Network.

1) *CTC token passing algorithm:* The CTC Token Passing algorithm, takes this sequence of letter probabilities, as well as a dictionary and a language model, as its input and computes a likely sequence of words. This algorithm works in output layer [2]. In the existing system, a bigram language model is used but in the proposed system, statistical N - gram language model will be use.

International Journal of Computer Science
ISSN:2348-6600
Oddity...Probe...Reviste...
Published by SK Research Group of Companies(SKRGC)
http://www.ijcsjournal.com Volume 3, Issue 1, No 2, 2015. ISSN: 2348-6600
Reference ID: IJCS-078 PAGE NO: 441-444.

Modification to Perform Keyword Spotting:
In order to find entire words but no sub words contained within longer words, add a "_" (white space) character to the front and the end of the keyword. The keyword to be spotted is preceded and succeeded by * to symbolize the any-node.

$$w' = \text{“}*\_l_1\ l_2 \ldots l_n\ \_*\text{”}.$$

If now use CTC Token Passing Algorithm for single word recognition to compute the probability of the word being *w'*. To receive a normalized value which can be threshold, take the logarithm of the probability $P_{CTC}(w|s)$ and divide it by the search words length:

$$f_{CTC}(w|s)= log(P_{CTC}(w|s)) / |w|.$$

VI. Conclusions

In this paper, we presented bidirectional long short-term neural networks in combination with a modified version of the Connectionist Token Passing algorithm. This system has several advantages compared to existing techniques. First, it is a line-based approach and does not need any word segmentation. Second, although the system needs to be trained, it does not require bounding boxes around characters or words, as is often needed in the keyword spotting literature. The only requirement is a transcription of the text lines in the training set. Finally, being derived from a general neural network-based handwritten text recognition system, any arbitrary string can be searched for, not just the words appearing in the training set.

Acknowledgment

Tejashri Chougule Author thanks Ms. A. N. Mulla, project guide, Department of Computer Science & Technology, Kolhapur, India, for her precious guidance, encouragement and continuous monitoring throughout the presented work.

References

[1] Volkmar Frinken, Andreas Fischer, R. Manmatha and Horst Bunke, "A Novel Word Spotting Method Based on Recurrent Neural Networks", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 33, no. 4 feb, 2012.

[2] Alex Graves, Marcus Liwicki, Santiago Fernandez, Roman Bertolami, Horst Bunke, and Jurgen Schmidhuber, "A Novel Connectionist System for Unconstrained Handwriting Recognition" , IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31,no. 5,feb 2009

[3] A. Vinciarelli, "A Survey on Off-Line Cursive Word Recognition", Pattern Recognition, vol. 35, no. 7, pp. 1433-1446, 2002.

[4] R. Plamondon and S.N. Srihari, "On-Line and O-Line Hand-writing Recognition: A Comprehensive Survey", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 1, pp. 63-84,Jan. 2000.

[5] S.-S. Kuo and O.E. Agazzi, "Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 842-848, Aug. 1994.

[6] R. Manmatha, C. Han, and E. Riseman, "Word Spotting : A New Approach to Indexing Handwriting" , Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, pp.631-637, 1996.

[7] U.-V. Marti and H. Bunke, "Using a Statistical Language Model to Improve the Performance of an HMM-Based Cursive Hand-writing Recognition System," Int'l J. Pattern Recognition and Artificial Intelligence, vol. 15, pp. 65-90, 2001.

[8] A. Bhardwaj, D. Jose, and V. Govindaraju, "Script Independent Word Spotting in Multilingual Documents" , Proc. Second Int'l Workshop Cross Lingual Information Access,pp. 48-54, 2008.

[9] Y. Leydier, F. Lebourgeois, and H. Emptoz , "Text Search for Medieval Manuscript Images" , Pattern Recognition, vol. 40, pp. 3552-3567, 2007.

[10] A. Kotcz, J. Alspector, M.F. Augusteijn, R. Carlson, and G.V. Popescu, "A Line-Oriented Approach to Word Spotting in Handwritten Documents" , Pattern Analysis and Applications, vol. 3, pp. 153-168, 2000.

[11] H. Cao, A. Bhardwaj, and V. Govindaraju, "A Probabilistic Method for Keyword Retrieval in Handwritten Document Images" , Pattern Recognition, vol. 42, no. 12, pp 3374-3382, http://dx.doi.org/10.1016/j.patcog.2009.02.003, Dec. 2009.

[12] J. Chan, C. Ziftci, and D. Forsyth, "Searching Off-Line Arabic Documents" , Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1455-1462, 2006.

[13] E. Saykol, A.K. Sinop, U. Güdükbay, O. Ulusoy, and A.E. Cetin, "Content-Based Retrieval of Historical Ottoman Documents Stored as Textual Images" , IEEE Trans. Image Processing, vol. 13, no. 3, pp. 314-325, Mar. 2004.

[14] R.F. Moghaddam and M. Cheriet, "Application on Multi-Level Classifier and Clustering for Automatic Word Spotting in Historical Document Images" , Proc. 10th Int'l Conf. Document Analysis and Recognition, vol. 2, pp. 511-515, 2009.